DOCUMENT RESUME

ED 442 847 TM 031 267

AUTHOR	Chang, Te-Sheng
TITLE	Student Ratings: What Are Teacher College Students Telling Us about Them?
PUB DATE	2000-04-00
NOTE	33p.; Paper presented at the Annual Meeting of the American Educational Research Association (New Orleans, LA, April 24-28, 2000).
PUB TYPE	Reports - Research (143) -- Speeches/Meeting Papers (150)
EDRS PRICE	MF01/PC02 Plus Postage.
DESCRIPTORS	Correlation; Educational Environment; Foreign Countries; Higher Education; Preservice Teachers; *Student Evaluation of Teacher Performance; Teacher Characteristics; *Teacher Education; *Undergraduate Students
IDENTIFIERS	*National Hualien Teachers College (Taiwan)

ABSTRACT

The purpose of this study was to investigate the effect of course, class, student, and instructor characteristics on student ratings of instruction. The sample included 437 undergraduate courses with 96 (21.2%) freshman courses, 140 (32.0%) sophomore courses, 119 (27.2%) junior courses, and 82 (18.8%) senior courses at National Hualien Teachers College, Taiwan, in the fall semester of the 1996-1997 academic year. The Student Ratings of Instruction form was used to measure student ratings of faculty performance. The results indicate that student enthusiasm, expected grade, and teacher grading standard are positively correlated with all five evaluation scores. Course difficulty is negatively corrected with all evaluation scores. The high semi-partial correlation coefficients consistently appear in course difficulty, student enthusiasm, and grading standard with all five evaluation scores. None of these 75 correlation coefficients show substantial nonlinearity. The percentages of variance explained by different combinations of background variables are 49.8%, 61.6%, 71.0%, 61.3%, and 66.8% for preparation, coverage, skills, assessment, and overall, respectively. Four of the background variables consistently appearing in the final regression equations are student enthusiasm, course difficulty, grading standard, and expected grade. (Contains 9 tables and 38 references.) (Author/SLD)

Reproductions supplied by EDRS are the best that can be made
from the original document.

ED 442 847

Running head: STUDENT RATINGS OF INSTRUCTION

Student Ratings: What Are Teacher College Students Telling Us about Them?

Te-Sheng Chang
National Hualien Teachers College
Hualien, Taiwan, 970, R.O.C.
achang@sparc2.nhltc.edu.tw

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☐ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

T.S. Chang

TM031267

This article was presented at the annual meeting of American Educational Research Association, held in New Orleans, LA, April 24-28, 2000.

Correspondence concerning this article should be addressed to Te-Sheng Chang, P.O. Box 20-76, Hualien, Taiwan, 970, R.O.C. To Contact by E-mail, please mail to achang@sparc2.nhltc.edu.tw

BEST COPY AVAILABLE

Abstract

The purpose of this study was to investigate the effect of course, class, student, and instructor characteristics on student ratings of instruction. The sample included 437 undergraduate courses with 96(21.2%) freshman courses, 140(32.0%) sophomore courses, 119(27.2%) junior courses, and 82(18.8%) senior courses at National Hualien Teachers College in the fall semester of the 1996-1997 academic year. The Student Ratings of Instruction (SRI) form was used to measure students' perceptions of faculty performance. The results indicate that student enthusiasm, expected grade and teacher grading standard are positively correlated with all five evaluation scores. Course difficulty is negatively correlated with all evaluation scores. The high semi-partial correlation coefficients consistently appear in course difficulty, student enthusiasm, and grading standard with all five evaluation scores. None of these 75 correlation coefficients show substantial non-linearity. The percentage of variance explained by different combinations of background variables are 49.8%, 61.6%, 71.0%, 61.3%, and 66.8% for preparation, coverage, skills, assessment, and overall, respectively. Four of the background variables consistently appearing in the final regression equations are student enthusiasm, course difficulty, grading standard, and expected grade.

Key words: Student Ratings, Course Evaluation, Faculty Evaluation, College Teaching

Teachers College

Student Ratings: What Are Teacher College Students Telling Us about Them?

Introduction

Background

The evaluation of course or teacher effectiveness has become more important than ever in the higher education sector (Wagner, 1999). One of the popular approaches to evaluate course or teacher effectiveness is student ratings of instruction. Student ratings of instruction have been in existence for seven decades since 1920s (Wachtel, 1998). The measurement of students' perceptions of instructors performance has been increasingly used as a major component in teaching evaluation (Feldman,1997; Jirovec, Ramanathan, & Alvarez, 1998; Wachtel, 1998). Some realistic reasons for this are as follows. (1) Students are an obvious and convenient choice for raters. (2) They have closely and recently observed a number of teachers. (3) They uniquely know how students think and feel. (4) Students' frank reactions can be a beneficial aid in refining course structure and teaching styles. (5) Student ratings are more objective than many other approaches, such as administrator evaluation, peer evaluation, self-rating, and classroom visitations (Arreola, 1995; Peterson, 1995). In addition, student ratings may serve three major functions: (1) aiding administrative evaluations of teaching effectiveness for decisions concerning pay increases, promotion, and tenure; (2) providing feedback to teachers for the purpose of improving instruction; and (3) helping students for courses or instructors selection (Centra, 1993; Cohen, 1980; Marsh & Roche, 1993).

Unlike most universities in western countries which have a long history in applying student ratings to course effectiveness, the program of student ratings has been employed at National Hualien Teachers College in Taiwan since the fall semester of 1996-1997 academic year. The faculty evaluation committee has decided that the result of student ratings is considered as a major component of faculty teaching performance. Despite the benefits and rationales offered to justify use of student ratings, many faculty members have hesitatingly warmed to the concept. Faculty contend that if variables can positively or negatively influence evaluations, then the influence of these variables must be taken into account before deciding to use such ratings in faculty advancement decisions. A crucial question here is what are the "important" background variables related to student ratings in teachers college. Are these "important" background variables different from those found in the western schools?

Variables Thought to Influence Student Ratings

A wealth of research exists in the area of student ratings, ranging from analyses of validity and reliability to studies parceling effects related to course, student, and teacher characteristics. This section provides an overview of the findings related to the variables which could conceivably exert an influence on student ratings scores.

Course Characteristics. Researchers reported that teachers of elective or non-required courses received higher ratings than teachers of required courses; a small to moderate positive relationship was found between course electivity and evaluation scores (Scherr & Scherr, 1990). This might be due to lower prior subject interest in required versus non-required courses. Most studies found that higher level courses tend to receive higher ratings (Chang, 1997; Marsh, 1987). Chang explained that this might be due to higher learning

enthusiasm in high level courses. Feldman (1978) reported that the association between course level and ratings is decreased when other background variables such as class size, expected grade, and electivity are controlled.

Greenwald and Gillmore (1998) reported that the introduction of mandatory student ratings led faculty to reduce course workloads and to make examinations easy in order to receive higher evaluation scores. They examined student ratings of hundreds of courses at University of Washington and found that professors who are easy graders receive better evaluations than do professors who are tougher. Marsh (1980) and Franklin, Thell, and Ludlow (1991), on the other hands, found a positive effect of course difficulty where more difficult courses were rated higher than less difficult courses. Wachtel (1998) argued that course level and student age might be confounding factors in more difficult courses.

Studies examining class size have arrived at various conclusions. Most researchers found that smaller classes tend to receive higher ratings (Mckeachie, 1990). Marsh and Dunkin (1992) argued that the class size effect is specific to certain dimensions of effective teaching, namely group interaction and instructional rapport. Another hypothesis was that the relationship between class size and student ratings is a U-shaped or curvilinear relationship, with small and large classes receiving higher ratings than medium-sized ones (Feldman, 1984). Some explanations which have been offered for this relationship included: departments may assign known superior teachers to large lecture classes or superior teachers may attract more students to their classes by virtue of their reputation (Wachtel, 1998).

Student Characteristics. Evidence suggested that students with greater interest in the subject area prior to the course tend to give more favorable teacher ratings (Prave & Bairl,

1993). Marsh and Dunkin (1992) asserted that the influence of prior interest on student ratings does not constitute a bias. They admitted that when ratings are used for summative purpose, the influence of student interest toward a subject can be a source of unfairness in that, but it is a function of the course and not the teacher.

The effect of a student's expected grade in a course on the student ratings has been one of the most controversial topics. Numerous authors argued in favor of the leniency hypothesis (Koshland, 1991; Nimmer & Stone, 1991) and against it (Marsh, 1987; Theall & Franklin, 1991). However, at this time, the consensus was definitely that there is a moderate positive correlation between expected grade and student ratings (Braskamp & Ory, 1994; Marsh, 1987; Marsh & Dunkin, 1992). The controversy concerned the interpretation of this association. Chacko (1983) showed that more strict grading standards led students to rate the instructor lower even on components of instruction unrelated to grading fairness, such as humor, self-reliance, and attitude toward students. Marsh (1987) suggested three plausible interpretations: (1) the leniency hypothesis, instructors with more lenient grading standards receive more favorable ratings; (2) the validity hypothesis, more effective instructors cause students to work harder, learn more and earn better grade; (3) the student characteristics hypothesis, pre-existing student characteristics such as prior subject interest affect both teaching effectiveness and student ratings.

The effect of student gender on student ratings is another controversial topic. Many studies reported that there was essentially no difference in ratings by male and female students, but a few have also come to a different conclusion (Watchel, 1998). Tatro (1995), for example, found that female students gave higher ratings than males. However, Koushki and Kuhn (1982) found the opposite results. In addition, some studies reported a tendency

for student to rate same-sex instructors slightly higher than opposite-sex instructors (Centra, 1993; Feldman, 1993).

Teacher Characteristics. Research typically indicated a positive effect of teacher rank on student ratings but a negative effect for age of the faculty member and years of teaching on ratings (Feldman, 1983). Feldman noted that while higher faculty rank is typically associated with higher overall ratings, the relationship can disappear or reverse when particular dimensions of teaching are examined. Discussion of the effect of teacher gender on student ratings appeared to be quite varied. In a two-part meta-analysis, Feldman (1992, 1993) reviewed existing research on student ratings of male and female teachers in both the laboratory and the classroom setting. In his review of laboratory studies, Feldman (1992) reported that the majority of studies reviewed showed no difference in the global evaluations of male and female teachers. In the minority of studies, in which difference was found, male instructors received higher overall ratings than females. Subsequently, in his review of classroom studies, Feldman (1993) again reported that the majority of studies reported no significant differences between the genders.

Grading standard perhaps generates the most suspicion about the validity of student ratings. Bridgeman (1986) and Owie(1985) compared summary evaluation scores of three groups, those receiving grades worse than expected, same as expected, and better than expected. Both of them found significant differences among the groups. The lowest evaluations came from the negative discrepancy group; the highest came from the zero discrepancy group for Bridgeman and the positive discrepancy group for Owie. Greenwald and Gillmore (1998) found that teachers can raise their ratings substantially by grading more leniently. They believed lenient grading leads to increased student ratings and is

easily performed. On the other hand, Roche and Marsh (1998) disagreed that lenient grading inflates student ratings.

Comprehensive Set. Research has confirmed the relationship between a comprehensive set of background variables and student evaluation of college teaching. Price and Magoon (1971) found that 11 background variables explained over 20 percent of the variance in 24 rating items. Similarity, Pohlmann (1975) reported that nine background variables explained over 20 percent of the variance in five evaluation items. In his study, course difficulty was the rating item best predicted, and it was correlated to a conceptually similar item concerning the hours spent outside of class. Brown (1976) showed that 11 background variables explained 14 percent of the variance in an average of student evaluation items, but indicated that grade accounted for the most variance. Burton (1975) found that eight background variables explained between 8 and 15 percent of the variable in instructor ratings over a seven-semester period of time, but indicated that the most important variable was student enthusiasm for the subject (cited in Marsh, 1980). Marsh (1980) found that 16 background variables explained between 0 and 25 percent of the variance in 11 evaluation scores and indicated the most influential variables were prior subject interest, expected grades, workload, and reason for taking the course.

It must be noted here that the mere existence of a correlation between a background variable and rating scores does not necessarily constitute a bias or a threat to the validity of student ratings (Brandenburg, Slinde, & Batista, 1977; Tatro, 1995). For example, if the student expected grade in the course is found to be associated with the rating which that student gives to the instructor, it does not necessarily follow that an instructor can obtain higher ratings merely by giving higher grades. Alternative explanations include the

possibility that more effective teaching will inspire students to work harder and earn better grades (Watchel, 1998).

Purpose

While research studies have indicated the different important variables influencing the student ratings of instruction in the west, it is significant to develop a similar study in the same field in colleges in eastern countries, especially in teachers college. Both faculty and students in teachers college are more conservative about student ratings than those in other schools. It is interesting to look into what are the important background variables related to student ratings in teachers college in Taiwan and to compare the result from this study to those from the western studies.

The purpose of this study was to investigate the relationship between student ratings of courses and a broad set of student, course, and instructor characteristics. The research was conducted with a reliable survey instrument that measured distinct evaluation factors, and was based on the ratings from a large number of classes. A stepwise regression analysis was used to determine the combined effect of the entire set of background variables, which of the set were the most important, and how this effect varied for the different dimensions of the student evaluations.

Definition of terms

Evaluation scores. Five student evaluation scores, used as dependent variables in this study, were described in the following:

1.preparation: An evaluation factor score representing three items [Item 1 to Item 3 from the Student Ratings of Instruction (SRI) form] for the extent to which students perceived the instructor to prepare for the class, to give the clear syllabus, and to

teach the class on time.

2 coverage: An evaluation factor score representing three items (Item 4 to Item 6 from SRI) for the degree to which student felt the instructor to present valuable materials to the subject and to emphasize analytic ability and conceptual understandings.

3 skills: An evaluation factor score representing four items (Item 7 to Item 10 from SRI) for the extent to which student perceived the instructor to encourage class discussions, to invite students to share their ideas, and to display enthusiasm, energy, and an ability to hold students' interest.

4 assessment: An evaluation factor score representing three items (Item 11 to Item 13 from SRI) for the extent to which student felt the value and fairness of graded materials in the course, and the value of assignments in adding appreciation and understanding of the subject.

5.overall instructor: A composite of the average of Item 1 to Item 13 from SRI.

Background variables. Information about course, class, student, and instructor characteristics was obtained on 15 different variables.

1. course difficulty: Student perceptions of the relative difficulty required by the course. An evaluation item score ranges from 1 to 5, 1:very easy; 3:medium; 5:very difficult.

2. course level: There are four levels for the course division, 1 for freshman, 2 for sophomore, 3 junior, and 4 for senior division.

3. electivity of course: The reason for students to take the course as elective (assigned as 1) or requirement (assigned as 0).

4. concentration: Students take the course as their major (assigned as 1) or not (assigned as 0) For example, the mathematics students in a mathematics course.

5. class size: The number of students are enrolled in the class.

6. proportion of male students in the class: The percentage of male students in the class.

7. enthusiasm toward the subject: Level of student enthusiasm for the subject or course. An evaluation item score ranges from 1 to 5, 1:very low; 3:medium; 5:very high.

8. student participation: Frequency of student participation into the class for the semester. An evaluation item score ranges from 1 to 5, 1:seldom; 3:medium; 5:always.

9. expected grade: The final grade students expected the instructor would give to them. An evaluation item score ranges from 1 to 5, 1:below 60; 2:60 to 69; 3:70 to 79; 4:80 to 89; 5: above 90.

10.teacher gender: 1:male instructor; 0:female instructor.

11. teacher rank: 1:full professor; 2:associate professor; 3: lecturer.

12.teacher status: 1: full-time teacher; 0: part-time teacher

13. teacher age: Instructor age was computed as with the formula of the year instructor born subtracted from 1997. (e.g., A teacher was born in 1961, he would be 36 years old in this study).

14 teacher degree: 1: bachelor, 2: master, 3: doctor.

15. grading standard: The discrepancy between student-expected grade and the grade students thought their teachers would give to them. Positive discrepancy means

grading standard is strict; while negative discrepancy means teacher grading standard is lenient. That is, the higher the discrepancy, the stricter the grading standard.

Method

Sample

The data for this investigation came from student course evaluations at National Hualien Teachers College and were drawn from the fall semester of the 1996-1997 academic year. Evaluations on which students failed to respond to questions that are key variables in the model were eliminated. Graduate courses and courses with enrollment less than five were also eliminated in this study. The final analytic sample included 12032 evaluations for 437 undergraduate courses offered in seven departments. These 437 courses consisted of 96(21.2%) freshman classes, 140(32.0%) sophomore classes, 119(27.2%) junior classes, and 82(18.8%) senior classes. It was possible that one instructor was rated by several courses and that one student contributed several ratings to the database. Given the sample size, it was expected that the effects of these repeated observations would be negligible. Table 1 describes the courses have constituted the study's sample.

Measures

The Student Ratings of Instruction (SRI) form used to measure students' perceptions of faculty performance was developed by the Academic Office of National Hualien Teachers College. It was composed of 13 questions rated on a 5-point Likert scale ranging from strongly agree (5-point) to strongly disagree (1-point). It yielded four dimensions: preparation (Item 1 to Item 3), coverage (Item 4 to Item 6), skills (Item 7 to Item 10), and

assessment (Item 11 to Item13). The average score of these thirteen items was treated as an overall score for the instructor. The α coefficients of internal consistency reliability from pilot study were .857, .917, .933, .927, and .969 for preparation, coverage, skills, assessment, and overall, respectively. These coefficients confirmed that the questionnaire was a reliable instrument. Principal components analysis was applied for each factor separately. Factor loadings for items designed to measure each factor were consistently large, between .885 and .944. Each principal components analysis had only one eigenvalue greater than one which indicated the items were pure indicators for their own factors. The four factors accounted for 87% of the total variance.

Procedures

To maximize respondent participation, the evaluation forms were distributed by the Vice-President of the Academic Administrative Office and administered by the head of class during the last two weeks of the semester. The instructor of record was absent from the classroom during the evaluation process. Anonymity of responses was emphasized. When the forms were completed they were returned to the administrator, placed in an envelope and sealed, prior to returning them to the Academic Administrative Office. The data were entered into the SPSS (Statistical Package for Social Science) analysis system.

Analytic Strategy

All analyses were performed on class-average responses for the sample. Fifteen background characteristics obtained from the survey and school database were course difficulty, course level, electivity, concentration, class size, proportion of male student in the class, student enthusiasm toward subject, participation, expected grade, teacher gender,

rank, age, degree, grading standard. Each of the 15 background variables was correlated with five student rating scores (preparation, coverage, skills, assessment, and overall). When sample size is large, it is easy for a correlation coefficient to be statistically significant but little practical significance. Therefore, attention was focused on those relationships which account for at least 5 percent ($r \geq .23$) of variance in any one of the evaluation scores (Marsh, 1980).

Quadratic and cubic components of each background variable were then tested to determine if any substantial non-linearity existed in the bivariate relationship. A relationship was considered to have substantial nonlinearity if quadratic and/or cubic components added at least 1 percent to the variance accounted for by the linear relationship and the total variance predicted by all components was at least 5 percent (Marsh, 1980).

Stepwise multiple regression was used to determine the combined effect of the background variables on each evaluation score and to determine which of the background variables consistently makes the largest contribution. Semipartial correlation coefficients for each background variable were computed to determine their uniqueness for the proportion of variance explained. This was accomplished by computing the proportion of variance that can be predicted by all but one of the background variables, and then determining the additional variance (the change in multiple R^2) that can be explained by the addition of the remaining variable. To simply the interpretations, a conservative criterion was used: An additional variable was included only if it added at least 1 percent to the variance which had already been explained by the previous set of variables (Marsh, 1980).

Results

Descriptive Statistics

Table 2 presents the mean, standard deviation, and range for each factor and overall evaluation scores across all 437 courses. Favorable student ratings are apparent, mean scores for preparation, coverage, skills, assessment, and overall are 3.97, 4.10, 3.88, 3.92, and 3.97, respectively. Like Amstrong's (1998) statement, most teachers are rated above average (about 4 on a 5-point scale). Obviously, the finding here is another evidence.

Correlation

Table 3 shows the zero-order correlation and semi-partial correlation coefficients between each of the 15 background variables and the five evaluation scores. Of the 75 zero-order correlation coefficients, 46 are statistically significant ($p < .05$), and only 27 of these 46 account for at least 5 percent of the variance ($r \geq .23$). Among these 27 correlation coefficients, student enthusiasm, participation, expected grade, and teacher grading standard are positively correlated with all evaluation scores. Student ratings tend to be higher when student enthusiasm is higher, when student participation is higher, and when expected grade is higher. This finding is similar to the studies by Burton (1975), Marsh (1980), and Braskamp and Ory (1994). Course level is positively correlated with coverage, skills, and overall evaluation scores. Student ratings in coverage, skills, and overall scores tend to be higher when course level is higher. This finding is same as the finding by Chang (1997) and Marsh (1987). Chang explained that this might be due to higher learning enthusiasm in higher level courses. Course difficulty is negatively correlated with all

evaluation scores. As Greenwald and Gillmore's (1998) study, faculty might reduce course difficulty to receive higher evaluation scores.

None of these 75 correlation coefficients show substantial non-linearity. Although 27 relationships between background variables and evaluation scores account for 5 percent of the variance, only 18 semi-partial correlation coefficients are greater than .10. The high semi-partial correlation coefficients consistently appear between course difficulty, student enthusiasm, and grading standard and all evaluation scores. That is, only course difficulty, student enthusiasm, and grading standard uniquely explain at least 1% of the variance in even each of the five evaluation scores. The semi-partial correlations between student participation and evaluation scores are not statistically significant. The expected grade does not have significant semi-partial correlation with evaluation scores but coverage.

Stepwise Regression

Table 4 through Table 8 present the summary of stepwise regression analysis for background variables predicting each of five evaluation scores. The number of background variables maintained in the final model are different. It is 5 variables for preparation, 7 for coverage, 8 for skills, 7 for assessment, and 6 for overall evaluation. The percentage of variance explained by different combinations of background variables are 49.8%, 61.6%, 71.0%, 61.3%, and 66.8% for preparation, coverage, skills, assessment, and overall, respectively.

Although more than five background variables are maintained in the different final regression models, the attention is paid to the variables only if the change in total variance accounted for from the step is greater than .01(1%). Based on this criterion, only three background variables consistently appear in the final regression equations through the

evaluation scores: student enthusiasm, course difficulty, and grading standard. Expected grade also appears in the most of final regression equations but preparation.

Table 9 includes the importance of the background variable for each regression model in terms of the magnitude of beta weights. As Table 9 shows, student enthusiasm toward subject is the most important variable in contribution to student ratings of instruction. Course difficulty, grading standard, and expected grade stand for the second, third, and fourth place, respectively. Like Marsh's (1980) study, student interest is the first important variable predicting student evaluation scores. Student interest into a course can be considered as a part of his enthusiasm toward the subject.

The course difficulty also has a strong contribution to the variance of the student evaluation scores. However, Marsh (1980) found the course difficulty has positive impact on student ratings, which is opposite from the finding by this study.

Discussion

The main issues considered in this study are whether teachers college students' evaluations of instructor effectiveness are affected by course, student, and teacher characteristics. The results of this study indicate two points: First, the most important variables related to student ratings of instruction are student enthusiasm, course difficulty, grading standard, and expected grade. Except for grading standard, the other three variables were also found by Marsh (1980). Student enthusiasm is tended to be the most important variable, which accounts for the most variance of student ratings. Marsh did not involve grading standard in his study, however, it is found to be one of the rating items best predicted in this study. Secondly, course difficulty, grading standard, and expected grade are all related to students' course grades. Taiwanese students, especially students in

teachers college pay more attention to their course grades than students in universities. Grading perhaps generates the most suspicion or importance about the validity of student ratings in teachers college in Taiwan.

This study provides a baseline for investigating the relationship between some course, student, and instructor background variables and student ratings of instruction in teachers college in Taiwan. It suggests that teaching-unrelated variables like course level, electivity, concentration, enrollment, student gender, teacher gender, rank, age, degree, and status which faculty thought might have influence on student ratings do not seem to make much difference in all evaluation scores (less than 1% of explanation). This increases confidence in the continued use of student ratings and simplifies their interpretation.

Like other western universities, student enthusiasm is the most important variable contributing to the student ratings in teachers college in Taiwan. In other words, how to encourage student's enthusiasm to learn should be the first thing for an instructor to do. Course difficulty is negatively related to student ratings, indicating that students who feel more capable in learning the subject rating their instructors more favorably. In short, faculty say they are more effective teachers when students are more motivated and more capable, and this effect is accurately reflected in the student ratings. Grading standard and expected grade are positively related to student rating, reflecting a possible unknown combination of easy grading and better student learning. This leaves another student ratings issue for future study.

Acknowledgments

The data in this study was provided by the Academic Administrative Office of National Hualien Teachers College.
I express my appreciation to the three AERA reviewers for comments on this paper.

References

Amstrong, J. S. (1998). Are student ratings of instruction useful? American Psychologist, 53(11), 1223-1224.

Arreola, R. A. (1995). Developing a comprehensive faculty evaluation system. Bolton, MA: Anker Publishing Company, Inc.

Brandenburg, D. C., Slinde, J. A., & Batista, E. E. (1977). Student ratings of instruction: Validity and normative interpretations. Research in Higher Education, 7, 67-78.

Braskamp, L. A., & Ory, J. C. (1994). Assessing faculty work. San Francisco: Jossey-Bass.

Bridgeman, W. J. (1986). Student evaluation viewed as a group process factor. Journal of Psychology, 120, 183-190.

Brown, d. L. (1976). Faculty ratings and student grades: A university-wide multiple regression analysis. Journal of Educational Psychology, 68, 573-578.

Burton, D. (19750). Student ratings-an information source for decision-making. In R. G. Cope (Ed.) Information for decisions in post-secondary education: Proceedings of the 15th Annual Forum. St Louis, MI: Association for Institutional Research.

Centra, J. A. (1993). Reflective faculty evaluation. San Francisco: Jossey-Bass.

Chacko, T. I. (1983). Student ratings of instruction: A function of grading standards. Educational Research Quarterly, 8, 19-25.

Chang, T. (1997). Student ratings of instruction at teachers college in Taiwan. Hualien, Taiwan: Graduate Institute of Compulsory Education, National Hualien Teachers College. (ERIC Document Reproduction Service No. ED 414 798)

Cohen, P. A. (1980). Effectiveness of student-rating feedback for improving college instruction: A meta-analysis of findings. Research in Higher Education, 13(4), 321-341.

Feldman, K. A. (1978). Course characteristics and college students' ratings of their teachers: What we know and what we don't. Research in Higher Education, 9, 199-242.

Feldman, K. A. (1984). Class size and college students' evaluations of teachers and courses: A closer look. Research in Higher Education, 21, 45-116.

Feldman, K. A. (1992). College students' views of male and female college teachers: Part I –evidence from the social laboratory and experiments. Research in Higher Education, 33, 317-375.

Feldman, K. A. (1993). College students' views of male and female college teachers: Part II-Evidence from students' evaluation of their classroom teachers. Research in Higher Education, 34, 151-211.

Feldman, K. A. (1997). Identifying exemplary teachers and teaching: Evidence from student ratings. In R. P. Perry & J. C. Smart (Eds.). Effective teaching in higher education: Research and practice (pp. 368-395). Bronx, NY: Agathon.

Franklin, J. Thell, M. & Ludlow, L. (1991). Grade inflation and student ratings: A closer look. Paper presented at the annual meeting of the American Educational Research Association, Chicago.

Greenwald, A. G., & Gillmore, G. M. (1998). How useful are student ratings? Research to comments on the current issues section. American Psychologist, 53(11), 1228-1229

Jirovec, R. L., Ramanathan, C. S., & Alvarez, A. R. (1998). Course evaluations: What are social work students telling us about teaching effectiveness? Journal of Social Work Education, 24(2), 229-236.

Koshland, D. E. (1991). Teaching and research. Science, 251, 249.

Koushki, P. A., & Kuhn, A. J. (1982). How reliabbe are student evaluations of teachers? Engineering Education, 72, 362-367.

Marsh, H. W. (1980). The influence of student, course, and instructor characteristics in evaluations of university teaching. American Educational Research Journal, 17, 219-237.

Marsh, H. W. (1987). Students' evaluations of university teaching: Research findings, methodological issues, and directions for future research. International Journal of Educational Research, 11, 253-388.

Marsh, H. W., & Dunkin, M. J. (1992). Students' evaluations of university teaching: A multidimensional perspective. In J. C. Smart (Ed.), Higher education: Handbook of theory and research (pp. 143-233). New York: Agathon Press.

Marsh, H. W., & Roche, L. (1993). The use of students' evaluations and an individually structured intervention to enhance university teaching effectiveness. American Educational Research Journal, 30, 217-251.

Mckeachie, W. J. (1990). Research on college teaching: The historical background. Journal of Educational Psychology, 82, 189-200.

Nimmer, J. G., & Stone, E. F. (1991). Effects of grading practices and time of rating on student ratings of faculty performance and student learning. Research in Higher Education, 32, 195-215.

Owie, I. (1985). Incongruence between expected and obtained grades and students' ratings of the instructor. Journal of Instructional Psychology, 12, 196-397.

Peterson, K. D. (1995). Teacher evaluation: A comprehensive guide to new directions and practices. Thousand Oaks, CA: Corwin Press, Inc.

Pohlmann, J. T. (1975). A multivariate analysis of selected class characteristics and student ratings of instruction. Multivariate Behavioral Research, 10, 81-92.

Prave, R. S., & Baril, G. L., (1993). Instructor ratings: Controlling for bias from initial student interest. Journal of Education for Business, 68, 362-366.

Price, J. A., & Magoon, A. J. (1971). Predictors of college student ratings of instructors. Proceedings of the 79th Annual Convention of the American Psychological Association, 6, 523-524.

Roche, L. a., & Marsh, H. W. (1998). Workload, grades, and students' evaluations of teaching: Clear understanding sometimes requires more patient explanations. American Psychologist, 53(11), 1230-1231.

Scherr, F. C., & Scherr, S. S. (1990). Bias in student evaluations of teacher effectiveness. Journal of Education for Business, 65, 365-358.

Tatro, C. N. (1995). Gender effects on student evaluations of faculty. Journal of Research and Development in Education, 28, 169-173.

Theall, M., Franklin, J. (1991). Using student ratings for teaching improvement. In M. Theall & J. Franklin (Eds.). Effective practices for improving teaching: New directions for teaching and learning (pp. 83-96). San Francisco: Jossey-Bass.

Wachtel, H. K. (1998). Student evaluation of college effectiveness: a brief review. Assessment and Evaluation in Higher Education23(2), 191-211.

Wagner, Z. M. (1999). Using student journals of course evaluation. Assessment and Evaluation in Higher Education, 24(3), 261-272.

Table 1. Description of Course, Student, and Teacher Characteristics (N = 437)

Course level	Freshman 96(22.0%)		Sophomore 140(32.0%)		Junior 119(27.2%)		Senior 82(18.8%)		Total	
Course electivity	N	%	N	%	N	%	N	%	N	%
Required	75	17.2	41	9.4	36	14.0	61	8.2	213	48.7
Elective	21	4.8	99	22.7	46	13.3	58	10.5	224	51.3
Course concentration										
Major	30	6.9	72	16.5	62	18.1	79	14.2	243	55.6
Not major	66	15.1	68	15.6	20	9.2	40	4.6	194	44.4
Teacher gender										
Male	58	13.3	97	22.2	59	17.6	77	13.5	291	66.6
Female	38	8.7	43	9.8	23	9.6	42	5.3	146	33.4
Teacher rank										
Full professor	22	5.0	30	6.9	16	6.4	28	3.7	96	22.0
Associate professor	30	6.9	63	14.4	24	10.8	47	5.5	164	37.5
Lecturer	44	10.1	47	10.8	42	10.1	44	9.6	177	40.5
Teacher Degree										
Doctor	35	8.0	57	13.0	19	9.8	43	4.3	154	35.2
Master	47	10.8	59	13.5	47	14.6	64	10.8	217	49.7
Bachelor	14	3.2	24	5.5	16	2.7	12	3.7	66	15.1
Teacher Status										
Full-time	87	19.9	131	30.0	78	24.3	106	17.8	402	92.0
Part-time	9	2.1	9	2.1	4	3.0	13	1.0	35	8.0
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Course difficulty	2.24	0.26	2.12	0.26	2.04	0.32	1.82	0.31	2.07	0.32
Male % in the class	0.28	0.14	0.29	0.20	0.39	0.20	0.42	0.20	0.34	0.20
Class size	29.08	8.11	20.95	8.70	18.47	8.13	16.52	8.43	21.23	9.46
Student enthusiasm	3.85	0.32	3.99	0.30	4.04	0.32	4.31	0.33	4.03	0.35
Student participation	4.28	0.23	4.28	0.24	4.41	0.24	4.58	0.25	4.37	0.26
Student expected grade	3.50	0.30	3.83	0.25	3.99	0.24	4.21	0.28	3.87	0.36
Teacher age	42.04	9.88	41.82	8.19	43.78	9.43	43.84	8.71	42.77	9.03
Teacher Grading standard	-0.09	0.09	-0.05	0.09	-0.07	0.15	-0.08	0.12	-0.07	0.12

Table 2. The Result of Student Ratings of Instruction for all Courses (N = 437)

Evaluation Factor	Item	*M*	*SD*	Range[a]
Preparation	1-3	3.97	0.40	1.99 - 4.87
Coverage	4-6	4.10	0.39	2.50 - 4.90
Skills	7-10	3.88	0.45	2.34 - 4.90
Assessment	11-13	3.92	0.37	1.99 - 4.81
Overall	1-13	3.97	0.38	2.55 - 4.79

a 5-point scale: 1 = poor; 5 = excellent.

Table 3 Correlations between 15 background variables and five Evaluation scores (N=437)

Variables	Preparation		Coverage		Skills		Assessment		Overall	
Course	*r*	Semi	*r*	Semi	*r*	Semi	*r*	Semi	*R*	Semi
Difficulty	-.582**	**.129**	-.654**	**.217**	-.716**	**.240**	-.598**	**.132**	-.680**	**.193**
Level	.209**	.049	.228**	.010	.264**	.027	.193**	.091	.239**	.042
Electivity	.039	.000	.143**	**.113**	.136**	.067	.119*	.021	.116*	.054
Concentration	.081	.020	.067	.010	.117*	.013	.149*	.053	.109*	.020
Class										
Male %	.130**	.000	.128**	.051	.115*	.040	.101*	.016	.126**	.000
Size	-.021	.036	-.014	.072	-.089	.013	-.126**	.054	-.066	.019
Student										
Enthusiasm	.660**	**.311**	.708**	**.294**	.761**	**.320**	.712**	**.361**	.755**	**.339**
Participation	.452**	.034	.464**	.049	.463**	.020	.424**	.011	.479**	.013
Expected grade	.363**	.030	.367**	**.118**	.431**	.082	.351**	.065	.403**	.079
Instructor										
Gender	.017	.037	.067	.000	.026	.047	.009	.050	.032	.037
Rank	.045	.062	-.053	.033	-.021	.020	-.025	.008	-.014	.000
Age	-.052	.084	.006	**.104**	-.028	.090	-.064	.058	-.036	.077
Degree	-.063	.000	-.161**	.089	-.157**	.093	-.068	.028	-.121*	.055
Status	-.012	.000	-.026	.055	-.042	.014	-.045	.014	-.033	.012
Grading standard	.167**	**.105**	.237**	**.143**	.276**	**.163**	.311**	**.220**	.263**	**.166**

* $p < .05$ ** $p < .01$ Correlations which are underlined indicate background variables accounting for at least 5% of the variance in an evaluation score. Semipartial correlations in bold are greater than .10

Table 4 · Summary of Stepwise Regression Analysis for Background Variables Predicting Student Ratings of Instruction in Preparation (N = 406)

Step	Variable	B	$SE\ b$	β	R	R^2	ΔR^2	F
1	a = .795							
	Student enthusiasm	.786	.043	.670	.770	.449	.449	328.667***
2	a = 2.072							
	Student enthusiasm	.608	.063	.517	.684	.468	.019	177.335***
	Course difficulty	-.267	.069	-.207				
3	a = 2.016							
	Student enthusiasm	.615	.063	.523	.692	.478	.010	123.420***
	Course difficulty	-.241	.069	-.186				
	Grading standard	.375	.127	.108				
4	a =1.927							
	Student enthusiasm	.615	.062	.523	.699	.489	.011	95.825***
	Course difficulty	-.251	.069	-.194				
	Grading standard	.374	.126	.107				
	Teacher rank	.051	.019	.097				
5	a = 2.013							
	Student enthusiasm	.638	.062	.543	.705	.498	.009	79.217***
	Course difficulty	-.294	.070	-.228				
	Course level	-.042	.016	-.106				
	Teacher rank	.055	.019	.103				
	Grading standard	.348	.125	.100				

* $p < .05$ ** $p < .01$; a: intercept; ΔR^2 : the increment of R^2 .The values which are underlined indicate the increment of R^2 greater than 1%.

Table 5 · Summary of Stepwise Regression Analysis for Background Variables Predicting Student Ratings of Instruction in Coverage (N = 406)

Step	Variable	B	$SE\ b$	β	R	R^2	ΔR^2	F
1	a = .821							
	Student enthusiasm	.831	.040	.711	.711	.506	.506	413.628***
2	a = 2.640							
	Student enthusiasm	.557	.057	.487	.740	.548	.042	244.171***
	Course difficulty	-.381	.062	-.303				
3	a = 2.557							
	Student enthusiasm	.567	.055	.496	.758	.574	.026	180.635***
	Course difficulty	-.342	.061	-.272				
	Grading standard	.554	.111	.164				
4	a = 3.590							
	Student enthusiasm	.605	.062	.523	.773	.597	.023	148.414***
	Course difficulty	-.485	.067	-.386				
	Expected grade	-.230	.048	-.206				
	Grading standard	.538	.109	.159				
5	a = 3.829							
	Student enthusiasm	.594	.054	.520	.778	.605	.008	122.768***
	Course difficulty	-.514	.067	-.409				
	Expected grade	-.256	.049	-.229				
	Grading standard	.495	.108	.147				
	Electivity	.076	.025	-.096				
6	a = 3.551							
	Student enthusiasm	.601	.054	.526	.783	.612	.007	105.070***
	Course difficulty	-.493	.067	-.393				
	Expected grade	-.221	.050	-.198				
	Grading standard	.488	.018	.144				
	Electivity	.098	.027	-.123				
	Class size	.004	.001	.093				
7	a = 3.619							
	Student enthusiasm	.607	.053	.531	.785	.616	.004	91.379***
	Course difficulty	-.478	.067	-.380				
	Expected grade	-.229	.050	-.205				
	Grading standard	.484	.107	.143				
	Electivity	.099	.027	-.124				
	Class size	.003	.001	.083				
	Teacher degree	-.038	.018	-.066				

* p < .05 ** p < .01; a: intercept; ΔR^2 : the increment of R^2 . The values which are underlined indicate the increment of R^2 greater than 1%

Table 6 · Summary of Stepwise Regression Analysis for Background Variables Predicting Student Ratings of Instruction in Skills (N = 406)

Step	Variable	B	$SE\ b$	β	R	R^2	ΔR^2	F
1	a = -.098							
	Student enthusiasm	.985	.042	.761	.761	.580	.580	557.380***
2	a = 2.277							
	Student enthusiasm	.651	.058	.503	.797	.636	.056	351.394***
	Course difficulty	-.497	.063	-.350				
3	a = 2.164							
	Student enthusiasm	.665	.055	.514	.820	.673	.037	275.949***
	Course difficulty	-.444	.061	-.313				
	Grading standard	.751	.110	.196				
4	a =3.157							
	Student enthusiasm	.702	.054	.543	.830	.690	.017	222.644***
	Course difficulty	-.582	.066	-.409				
	Grading standard	.735	.108	.192				
	Expected grade	-.221	.048	-.175				
5	a = 3.368							
	Student enthusiasm	.692	.054	.535	.834	.695	.005	182.070***
	Course difficulty	-.607	.066	-.427				
	Expected grade	-.244	.049	-.193				
	Grading standard	.698	.108	.182				
	Electivity	.067	.025	-.074				
6	a = 3.411							
	Student enthusiasm	.699	.054	.541	.836	.699	.004	154.195***
	Course difficulty	-.587	.067	-.412				
	Expected grade	-.249	.048	-.197				
	Grading standard	.693	.107	.181				
	Electivity	.070	.026	-.078				
	Teacher degree	-.043	.019	-.064				
7	a = 3.791							
	Student enthusiasm	.672	.053	.520	.841	.707	.008	137.123***
	Course difficulty	-.610	.066	-.429				
	Expected grade	-.238	.048	-.188				
	Grading standard	.693	.106	.181				
	Electivity	.065	.025	-.072				
	Teacher degree	-.067	.020	-.101				
	Teacher age	-.005	.001	-.100				
8	a = 3.807							
	Student enthusiasm	.675	.053	.521	.842	.710	.003	121.349***
	Course difficulty	-.622	.066	-.438				
	Expected grade	-.232	.048	-.184				
	Grading standard	.698	.106	.182				
	Teacher degree	-.070	.020	-.105				
	Teacher age	-.004	.002	-.087				
	Electivity	.067	.025	-.075				
	Male % in the class	-.131	.066	-.058				

* $p < .05$ ** $p < .01$; a: intercept; ΔR^2 : the increment of R^2 . The values which are underlined indicate the increment of R^2 greater than 1%

Table 7 · Summary of Stepwise Regression Analysis for Background Variables Predicting Student Ratings of Instruction in Assessment (N = 406)

Step	Variable	B	$SE\ b$	β	R	R^2	ΔR^2	F
1	a = .800							
	Student enthusiasm	.773	.038	.712	.712	.508	.508	417.281***
2	a = .956							
	Student enthusiasm	.749	.035	.691	.758	.575	.067	272.529***
	Grading standard	.831	.105	.259				
3	a = .850							
	Student enthusiasm	.799	.038	.737	.765	.585	.010	189.097***
	Grading standard	.817	.103	.255				
	Course level	-.041	.013	-.112				
4	a =1.753							
	Student enthusiasm	.681	.051	.628	.773	.597	.012	148.399***
	Grading standard	.767	.103	.239				
	Course difficulty	-.196	.058	-.165				
	Course level	-.051	.013	-.140				
5	a = 1.747							
	Student enthusiasm	.694	.051	.640	.777	.604	.007	122.089***
	Grading standard	.769	.102	.240				
	Course difficulty	-.196	.057	-.165				
	Course level	-.051	.013	-.140				
	Teacher gender	-.069	.025	-.086				
6	a = 1.902							
	Student enthusiasm	.687	.051	.634	.780	.609	.005	103.638***
	Grading standard	.759	.102	.237				
	Course difficulty	-.212	.058	-.178				
	Course level	-.064	.014	-.176				
	Teacher gender	-.066	.025	-.083				
	Class size	-.003	.001	-.079				
7	a = 2.401							
	Student enthusiasm	.694	.051	.640	.783	.613	.004	90.227***
	Grading standard	.760	.102	.237				
	Course difficulty	-.267	.063	-.224				
	Course level	-.047	.016	-.129				
	Expected grade	-.117	.055	-.110				
	Class size	-.003	.001	-.087				
	Teacher gender	-.056	.026	-.070				

* $p < .05$ ** $p < .01$; a: intercept; ΔR^2 : the increment of R^2 . The values which are underlined indicate the increment of R^2 greater than 1%.

Table 8 、Summary of Stepwise Regression Analysis for Background Variables Predicting Student Ratings of Instruction in Overall (N = 406)

Step	Variable	B	$SE\ b$	β	R	R^2	ΔR^2	F
1	a = .580							
	Student enthusiasm	.840	.036	.757	.757	.574	.574	543.591***
2	a = .708							
	Student enthusiasm	.820	.034	.740	.785	.617	.043	325.144***
	Grading standard	.688	.101	.210				
3	a = 2.094							
	Student enthusiasm	.625	.049	.564	.802	.643	.026	241.709***
	Grading standard	.619	.099	.189				
	Course difficulty	-.293	.054	-.241				
4	a =2.962							
	Student enthusiasm	.657	.048	.593	.813	.660	.017	194.948***
	Course difficulty	-.413	.059	-.339				
	Grading standard	.606	.097	.185				
	Expected grade	-.193	.043	-.178				
5	a = 3.172							
	Student enthusiasm	.639	.049	.577	.815	.664	.004	158.411***
	Course difficulty	-.434	.060	-.356				
	Grading standard	.605	.096	.185				
	Expected grade	-.187	.043	-.172				
	Teacher age	-.003	.001	-.065				
6	a = 3.275							
	Student enthusiasm	.639	.049	.577	.817	.668	.004	133.620***
	Course difficulty	-.424	.060	-.348				
	Grading standard	.603	.096	.184				
	Expected grade	-.188	.043	-.173				
	Teacher age	-.004	.001	-.089				
	Teacher degree	-.036	.018-	-.063				

* p < .05 ** p < .01; a intercept; ΔR^2 : the increment of R^2 The values which are underlined indicate the increment of R^2 greater than 1%.

Table 9 · Summary of the Order of Selection for Background Variables in Stepwise Regression Analysis in Predicting Student Ratings of Instruction Scores

Variable	Preparation	Coverage	Skills	Assessment	Overall
Course					
Difficulty	2	2	2	3	2
Level	3			4	
Electivity		5	7		
Concentration					
Class					
Male %			8		
Size		6		6	
Student					
Enthusiasm	1	1	1	1	1
Participation					
Expected grade		3	3	5	4
Instructor					
Gender				7	
Rank	4				
Age			6		5
Degree		7	5		6
Status					
Grading standard	5	4	4	2	3

a: The numbers indicate the order of selection for background variables in stepwise regression analysis.